Distinguishing Heterogeneity From Decreasing Hazard Rates

Dean A. Follmann
Center for Naval Analyses
Alexandria, VA 22302

Matthew S. Goldberg
The Rand Corporation
Santa Monica, CA 90406

Different copies of a repairable machine, or units, often exhibit different failure rates. If this heterogeneity is ignored, a statistical model of the time until failure may estimate a spurious decreasing hazard rate, resulting in incorrect inferences concerning the risk following repair. This article develops a renewal-proces model that accommodates both heterogeneity across units and decreasing hazard rates. Failure times for each unit are assumed Weibull, and the Weibull scale parameter is assumed to vary across units according to a gamma distribution. The model is illustrated using both Proschan's air-conditioner data and data on a U.S. Navy radar.

KEY WORDS: Overdispersion; Reliability; Renewal process; Repairable systems; Repeated measures; Weibull distribution.

1. INTRODUCTION

Many machines or units are repaired rather than replaced after a failure. Often the analysis of such units ignores the fact that each machine produces several failure times and assumes that, possibly given some explanatory variables, failure times are independent and identically distributed (iid) across units. The purpose of this article is to provide a model that relaxes this assumption.

If failure times across units are assumed iid without justification, heterogeneity across units may incorrectly result in the conclusion that failure times have a decreasing hazard rate (Proschan 1963). Suppose that the lifetimes of each unit follow an exponential (λ) distribution, but that λ varies across units according to some nondegenerate probability distribution. Then the distribution of lifetimes across units will exhibit a decreasing hazard rate. Intuitively, units with the larger λ's tend to fail first, so the average hazard rate of the remaining units tends to decrease with time.

The distinction between a heterogeneous population of units and a true decreasing hazard rate can be critical. Consider the difference in the two views for describing the lifetime following repair. If, conditional on a unit, each lifetime follows an exponential distribution, then a new or just-repaired unit is no more likely to fail than an old unit. Therefore, there is no increased risk following repair. But, if each lifetime has a decreasing hazard rate, then repair is followed by a higher risk. In other words, a unit that has survived its initial "burn-in" period is less likely to fail than a new or just-repaired unit.

There are several possible approaches to discriminating between heterogeneity and decreasing hazard rates. If there is a small number of units each with a large number of failures, then it may be possible to estimate a separate survival model for each unit. The fitted models for the different units could then be compared to assess the amount of heterogeneity. A related approach would be to estimate a combined model for all of the units but to include dummy variables to identify the individual units. Unfortunately, both of these approaches fail if there is a large number of units or if most units have few failures.

This article develops and estimates a statistical model that accommodates both decreasing hazard rates and heterogeneity across units. Specifically, the lifetimes of each unit are assumed to follow a Weibull distribution, and the Weibull scale parameter is itself drawn from a gamma mixing distribution. The Weibull shape parameter indicates whether the hazard rate for each unit is decreasing, constant, or even increasing. The variance of the gamma mixing distribution indicates the degree of heterogeneity across the units.

The model developed here is illustrated using two different data sets. First, the model is used to support Proschan's (1963) analysis of the well-known air-conditioner data. Second, Kujawski and Rypka (1978) observed that equipment reliability is adversely affected by on–off cycling. The model is also applied to data on a U.S. Navy radar in an attempt to describe failures and to measure the effect of on–off cycling on radar reliability.

We work with repairable systems and assume that the sequence of failure times for each unit forms a renewal process. That is, we assume that times between successive failures of a unit are iid (given certain explanatory variables); hence the time index is reset to 0 after each repair. The nonhomogeneous Poisson process is the other important point process used to model repairable systems. The distinction between these two modeling approaches was emphasized by Ascher and Feingold (1984). Statistical techniques for estimating nonhomogeneous Poisson processes were discussed by Lee (1980a,b) and Lawless (1987). Choosing between a renewal process and a nonhomogeneous Poisson process is difficult and will be left for future research.

2. HETEROGENEITY OR DECREASING HAZARD

Given a collection of data on repairable units, a simple analysis might assume that times between failures are iid, possibly given some explanatory variables, and use a flexible parametric model to describe failure times. A popular model that allows a nonconstant hazard rate is the Weibull. The Weibull density is given by

$$f(t_i | \lambda, \gamma) = \lambda \gamma t_i^{\gamma - 1} \exp(-\lambda t_i^{\gamma}), \qquad t_i > 0. \quad (1)$$

It follows that the hazard rate is equal to

$$f(t_i | \lambda, \gamma)/[1 - F(t_i | \lambda, \gamma)] = \lambda \gamma t_i^{\gamma - 1}, \qquad t_i > 0. \quad (2)$$

The parameter $\lambda > 0$ is known as the scale parameter, whereas $\gamma > 0$ is known as the shape parameter. The hazard rate is constant if $\gamma = 1$, decreasing if $0 < \gamma < 1$, and increasing if $\gamma > 1$.

Suppose that the hazard rate is constant for each unit but that the units are heterogeneous. Then estimation of a Weibull model under the (false) iid assumption will lead to a spurious decreasing hazard. This result is made precise in the following theorem.

Theorem 1. Let $\{T_{ij}\} = T_{i1}, \ldots, T_{im_i}$ represent the times between failure for the ith repairable unit. Let the $\{T_{ij}\}$ have independent exponential (λ_i) distributions, conditional on a value of λ_i. Let λ_i have a nondegenerate distribution across units. Then the unconditional distribution of T_{ij} has a decreasing hazard rate.

This theorem was proven by Proschan (1963) for the case in which the hazard rate for a given unit is drawn from a discrete mixing distribution. The theorem as stated by Lawless (1982, pp. 49–50) for the case of a continuous mixing distribution. The proof in the continuous case follows from the Schwartz inequality for integrals.

It can be demonstrated, via simulation, that if a Weibull distribution is estimated from data generated by a gamma mixture of exponentials, the estimated Weibull has a decreasing hazard rate ($\hat{\gamma} < 1$). Furthermore, the departure from 1 varies directly with the coefficient of variation of the gamma distribution (results are available on request). Therefore, the greater the heterogeneity across units, the more likely one is to incorrectly estimate a decreasing hazard if times between failure are assumed iid across units. Even with heterogeneity, however, conditional hazard rates may be constant, increasing, or decreasing. A sensible model needs to allow for both effects, heterogeneity across units and a nonconstant hazard rate.

3. STATISTICAL MODEL

Data are assumed available on repairable units $i = 1, \ldots, n$. Each unit operates until failure, the failed components are repaired or replaced, and the unit begins to operate again until another failure, and so forth. Therefore, multiple lifetimes t_{ij} ($j = 1, \ldots, m_i$) are observed for the ith unit. It is possible that t_{ij} represents a censored observation. The recording may stop on a unit, or replacement may occur before failure. To identify this situation, define an indicator variable d_{ij} equal to 1 for a complete observation and 0 for a censored observation.

It is assumed that lifetimes for the ith unit are drawn from a Weibull distribution with scale parameter λ_i and shape parameter γ [Eq. (1)]. It is further assumed that the scale parameter λ_i represents a random draw from a gamma mixing distribution,

$$g(\lambda) = \frac{\beta^{\alpha}}{G(\alpha)} \lambda^{\alpha - 1} \exp(-\beta \lambda), \qquad \lambda > 0, \quad (3)$$

where $G(\ )$ denotes the gamma function. Note that the gamma distribution has mean $\mu = \alpha/\beta$ and variance $\sigma^2 = \alpha/\beta^2$. We must emphasize that a scale parameter λ_i is drawn only once per unit, not once per lifetime. The single draw of λ_i for the ith unit generates all of its lifetimes t_{ij} ($j = 1, \ldots, m_i$), and the lifetimes are assumed independent (conditional on λ_i). The gamma mixing distribution models the heterogeneity across units, whereas the Weibull shape parameter describes the hazard rate within each unit. The choice of a gamma is made largely for mathematical convenience. Manton, Stallard, and Vaupel (1986), however, indicated that inference may not be very sensitive to the postulated mixing distribution.

The Weibull distribution may be modified to include explanatory variables thought to affect unit lifetimes (see, e.g., Prentice 1973). For example, variables may be available that measure the age of a unit or other features thought to affect failure. Let x_{ij} denote a vector of explanatory variables corresponding to lifetime t_{ij}. Assume that the explanatory variables serve to multiply the hazard rate by a factor $\exp(x_{ij}\theta)$, where θ is a vector of unknown coefficients. This assumption of proportional hazard rates is quite common in the survival literature (see Cox and Oakes 1984; Kalbfleisch and Prentice 1980) and can be tested. Note that the x_{ij} vector does not include a leading one for an intercept term. The intercept corresponds to μ, the mean of the mixing distribution.

To develop the likelihood function for the sample, assume that censoring is noninformative [but see Lagakos (1979) for generalizations]; that is, conditional on a value of λ_i, a complete observation ($d_{ij} = 1$) contributes a term $f(t_{ij} \mid \lambda_i)$ to the likelihood function. A censored observation ($d_{ij} = 0$) contributes a term $1 - F(t_{ij} \mid \lambda_i)$, where $F(t_{ij} \mid \lambda_i)$ is the Weibull cumulative distribution function. Therefore, conditional on λ_i, the total contribution of the ith unit to the likelihood function is equal to

$$L_i(\lambda_i, \gamma, \theta) = \prod_{j=1}^{m_i} [f(t_{ij} \mid \lambda_i, \gamma, \theta)]^{d_{ij}} \times [1 - F(t_{ij} \mid \lambda_i, \gamma, \theta)]^{1-d_{ij}}. \quad (4)$$

The unconditional contribution of the ith unit is obtained by integrating out λ_i:

$$L_i(\alpha, \beta, \gamma, \theta) = \int_0^\infty L_i(\lambda, \gamma, \theta) g(\lambda)\, d\lambda, \quad (5)$$

where $g(\lambda)$ is the gamma mixing distribution in Equation (3). Finally, the total likelihood function for the sample equals the product of the contributions of all n units;

$$L(\alpha, \beta, \gamma, \theta) = \prod_{i=1}^{n} L_i(\alpha, \beta, \gamma, \theta) = \prod_{i=1}^{n} \int_0^\infty L_i(\lambda, \gamma, \theta) g(\lambda)\, d\lambda. \quad (6)$$

Substituting the Weibull and gamma distributions into Equation (6) and performing the integration yields

$$L(\alpha, \beta, \gamma, \theta) = \prod_{i=1}^{n} \frac{G(\alpha + d_{i+})}{G(\alpha)} \left(\frac{\beta}{\beta + t_{i+}^{(\gamma)}}\right)^{\alpha} \times \left(\frac{\gamma}{\beta + t_{i+}^{(\gamma)}}\right)^{d_{i+}} \exp\left(\theta \sum_{j=1}^{m_i} d_{ij} x_{ij}\right) \prod_{j=1}^{m_i} t_{ij}^{(\gamma-1)d_{ij}}, \quad (7)$$

where $d_{i+} = \sum_{j=1}^{m_i} d_{ij}$, $t_{i+}^{(\gamma)} = \sum_{j=1}^{m_i} t_{ij}^{\gamma} \exp(x_{ij}\theta)$, and $G(\ )$ again denotes the gamma function. Precisely the same likelihood function was obtained by Wild (1983) as his model 2. He, however, did not use this model to distinguish heterogeneity from decreasing hazard rates. Instead, he was primarily concerned with the lack of efficiency in estimates of regression parameters when the mixing distribution is unspecified. Crowder (1985) worked with a similar model, except that he restricted the units to have equal numbers of failure times ($m_i = m$ for all i). Finally, Johnson and Kotz (1972, pp. 288–289) discussed this distribution and referred to it as the multivariate Burr distribution. The identifiability of the parameters of the model follows from results of Elber and Ritter (1982) and Heckman and Singer (1984).

Several special cases are contained in (7). If $\gamma = 1$, (7) is the likelihood for a gamma mixture (over units) of exponentials, whereas if the gamma mixing distribution is degenerate, (7) reduces to a Weibull likelihood. Thus the two "extreme" causes of apparent decreasing hazard rates are allowed—solely due to heterogeneity or not at all due to heterogeneity. For completeness, we note that if units are not repairable ($m_i = 1$), (7) reduces to the model introduced by Dubey (1968) and also considered by Lancaster (1979) and Lancaster and Nickell (1980). If a further assumption is made that $\gamma = 1$, (7) reduces to a Pareto likelihood with explanatory variables.

4. ESTIMATION AND MODEL CHECKING

The logarithm of Equation (7), or log-likelihood, may be maximized numerically to yield estimates of the parameters μ, σ^2, γ, and θ. Newton's method may be used to set the gradient of the log-likelihood to 0, or one of the quasi-Newton variations (see, e.g., Dennis and Schnabel 1983) that approximate the Hessian of the log-likelihood may be more convenient. Initial guesses of μ, θ, and γ may be obtained from a Weibull regression. An initial guess of $\sigma = \mu$ corresponds to an exponential mixing distribution and has worked well for us.

The proportional-hazard-rate assumption may be examined using the procedure suggested by Kay (1977) (see also Cox and Oakes 1984, pp. 112–113; Kalbfleisch and Prentice 1980, pp. 89–98). For binary explanatory variables, Kay's approach was to dichotomize the sample in turn for each variable and to estimate two models, one for each subsample. The proportional-hazard-rate assumption is met if the two estimated hazard rates are proportional. Under the Weibull baseline hazard assumed in this article, proportionality will be reflected by equal values of γ in the two subsamples. Continuous explanatory variables can be tested in analogous fashion by using "low" versus "high" values to define the subsamples.

To assess the assumption of a gamma mixing distribution, one could estimate separate λ_i's for each unit using a single Weibull regression with dummy variables for each unit and then compare these estimates to a gamma distribution. There may be difficulties, however, if there is a large number of units, each with few failures, and it may be necessary to include only units with a "reasonable" number of failures. If there are only a few units with many failuees, the preceding technique may be sufficient to model the data.

One common technique to assess overall goodness

Table 1. Maximum Likelihood Estimates for the Proschan Data

Model	Parameters γ	$E(\lambda)$	$Var(\lambda)$	Log-likelihood
Exponential	—	.011 (.001)	—	−1,178.8
Weibull	.925 (.057)	.016 (.005)	—	−1,177.6
Gamma mixture of exponentials	—	.011 (.001)	1.2×10^{-5} 2.7×10^{-6}	−1,174.9
Gamma mixture of Weibulls	.975 (.020)	.012 (.001)	1.4×10^{-5} 3.0×10^{-6}	−1,174.8

NOTE: Asymptotic standard errors appear in parentheses.

of fit uses *generalized residuals* (see Cox and Oakes 1984, pp. 88–89; Kay 1977). The basic idea follows from the result that if a continuous random variable T has distribution function F, then $F(T)$ has a uniform distribution. Therefore, if F is estimated by $\hat{F}$, the collection of transformed data, $\hat{F}(t_1), \ldots, \hat{F}(t_n)$, should roughly behave like an iid sample from a uniform distribution. In practice, some observations will be right-censored and so will be the corresponding $\hat{F}(t_i)$. A Kaplan–Meier survival-curve estimate based on the transformed data should look like a uniform survival curve—that is, a line connecting (1, 0) and (0, 1) if the model fits. The generalized residuals can also be grouped according to explanatory variables and examined to assess their dependence on the explanatory variables.

Although this technique is not directly applicable when there is heterogeneity across units, it can be easily modified using a suggestion of Lawless (1987). Recall that, conditional on λ_i, $F(T_{i1} \mid \lambda_i), \ldots, F(T_{im_i} \mid \lambda_i)$ are iid uniform. Therefore, the preceding technique can be applied as long as λ_i is estimated in addition to the parameters. A simple estimate of λ_i is given by the posterior mean

$$E[\lambda_i \mid (t_{i1}, d_{i1}), \ldots, (t_{im_i}, d_{im_i})] = (d_{i+} + \alpha)/(t_{i+}^{(\gamma)} + \beta),$$

where α and β are parameters of the gamma mixing distribution.

5. PROSCHAN DATA

As our first example, we apply the proposed statistical model to the data reported by Proschan (1963). He collected data on service hours between failures for air-conditioning systems on 13 Boeing 720 jet aircraft. He did not formulate an explicit model. Instead, he used hypothesis-testing procedures to arrive at two conclusions. First, if the data on all 13 aircraft are pooled together, the distribution function of the pooled data exhibits a decreasing hazard rate. Second, each individual aircraft exhibits a constant hazard rate, although the constant hazards may vary across aircraft.

Proschan's (1963) hypothesis-testing procedures lead to the conclusion that hazard rates are constant within aircraft but heterogeneous across aircraft. These procedures do not seem flexible enough, however, to accommodate a situation in which hazard rates are both nonconstant within aircraft and heterogeneous across aircraft. By contrast, the statistical model developed in this article does allow both effects, the nonconstant hazard rate being measured by the parameter γ and heterogeneity being measured by σ^2.

Several other authors have analyzed Proschan's data. For example, Ascher and Feingold (1984) argued that these data may not be consistent with a renewal process. To illustrate the proposed model, however, we will maintain the renewal assumption. The results of estimating several nested models culminating in (7) appear in Table 1.

First, a simple exponential model was estimated. Next, the exponential model was generalized to a Weibull model, but heterogeneity across aircraft was still not allowed (i.e., σ^2 was constrained to 0). The estimated Weibull shape parameter, γ, is less than 1.0, giving some evidence of a decreasing hazard. The estimate of γ, however, lies within two standard errors of 1.0, and the improvement in log-likelihood is not significant.

Next, a gamma mixture (across units) of exponential distributions was estimated. Both the Wald test statistic (i.e., the parameter estimate divided by its estimated asymptotic standard error) and the likelihood ratio test imply that σ^2 is significantly larger than 0. Both test procedures are conservative under the null hypothesis, where σ^2 is on the boundary of the parameter space. Chernoff (1954) demonstrated that the likelihood ratio statistic has an asymptotic distribution that is chi-squared with probability .5 and is zero with probability .5. Moran (1971) demonstrated that the Wald statistic has an asymptotic distribution that is standard normal with probability .5 and is zero with probability .5. Therefore, the mixture of exponentials seems to explain the data better than the homogeneous Weibull model.

Finally, the full gamma mixture of Weibull distributions was estimated. The estimate of σ^2 is essentially unchanged from the previous case, but the estimate of γ is within two standard errors of 1.0. Moreover, the log-likelihood is virtually identical to that of the simpler mixture of exponentials. Therefore, it appears that the mixture of exponentials provides the best summarization of the data of the four models considered.

These results support Proschan's (1963) major findings. For other data sets in which both non-

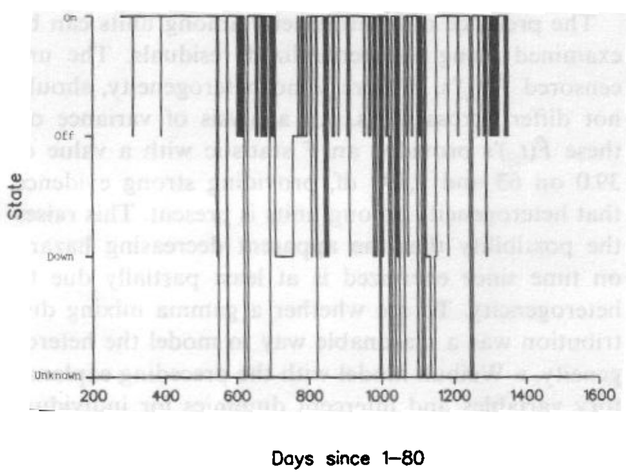

Figure 1. Usage History of a Single Radar. Horizontal line segments represent time spent in a state; vertical line segments represent transitions between states.

constant hazards and heterogeneity are present, however, the benefits of the statistical model will be much more pronounced. An example of such a data set is given in the next section.

6. NAVY DATA

Data were collected on a U.S. Navy radar for the period 1980–1985. Normal use of these radars involves intermittently turning the radar off and on for a variety of reasons, including operational concerns and preventive maintenance. Occasionally, the radars fail and are then repaired. Sixty-three radars were observed. There were 6,019 energized (on) periods, for an average of 96 per radar. Of these periods, 4,565 or 76% ended in censoring—either the machine was turned off or data recording stopped on a particular unit. Figure 1 provides an actual usage history for one of the radars. Thick solid black lines indicate that the radar switched frequently between different states.

Table 2. Sample Statistics and Explanatory Variables for Navy Data

Variable	Mean	Standard deviation
Days of observation per machine	1,030.4	444.9
Days energized per machine	397.0	217.3
Failures per machine	22.7	13.3
Days since energized	4.20	7.07
Days since repair	11.04	17.19
First "on" since repair	.33	.47
On–off cycles per day	.85	1.19
Calendar year	2.20	1.40
Cumulative failures	20.91	15.72

Several explanatory variables that were thought likely to affect failure were measured. Some summary statistics are given in Table 2. Note that the variables are measured at the start of each energized period.

The first three rows of Table 2 describe overall usage and reliability of the radars. Radars were observed for an average of almost three years, and about 40% of that time they were energized (the remaining time the radars were either off or broken). The radars averaged about 23 failures per machine. The fourth row of Table 2 indicates that failure or censoring occurred on average 4.2 days after being energized. The last five rows describe explanatory variables. The first three describe aspects of the "local" history of the unit, and the last two describe aspects of the "global" history. Specifically, the first measures operating time since repair—that is, the number of days the machine was operating from the end of the last repair until the start of the current "on" period. The second is set to 1 for the first "on" period following repair and is set to 0 otherwise. The third explanatory variable measures on–off cycles per day since the last repair. This covariate is set to 0 for the first "on" period following repair. The fourth represents calendar year and is coded as 0 for 1980, 1 for 1981, and so forth. The fifth measures cumulative failures since the beginning of the data set, January 1980.

Preliminary analysis indicated that the hazard rate decreased with both the time since repair and the time since a radar was energized (turned on). In effect, there are two possible time scales for the hazard function. The problem of multiple time scales was discussed by Farewell and Cox (1979) and Prentice, Williams, and Peterson (1981). The solution of Prentice et al. was essentially to use one of the time scales as the time index t in the model and enter the other time scale as an explanatory variable. Our approach is to use time since energized as the time scale and time since repair as a time-fixed explanatory variable. This formulation allows the hazard to vary as a function of both time since energized and time since repair.

Since time since energized is used as the time index, if there were no explanatory variables we would be assuming that each time a radar is turned on the process renews itself. This is a strong assumption. Use of the explanatory variables, especially time since repair, calendar year, and cumulative failures, can be viewed as ways of weakening this assumption.

Table 3 contains estimates of the Weibull regression model assuming no heterogeneity. Most of the explanatory variables are statistically significant. The units are more likely to fail during the first "on" period following repair, and generally failure is more likely the more recent a repair has been. Previous

Table 3. Maximum Likelihood Estimates for the Weibull Regression

Variable	Parameter	Standard error
μ	.114	.010
σ^2	—	—
γ	.764	.023
Days since repair	−.0059	.0020
First "on" since repair	.216	.078
On–off cycles per day	.039	.035
Calendar year	−.217	.033
Cumulative failures	.011	.003
Log-likelihood	−5,442.1	

on–off cycling does not appear to be important. There is a downward trend in the hazard rate over the sample period, 1980–1985; however, the more failures a particular machine has had, the more likely it is to fail again. Of particular interest is the estimate of γ. Its value of .764 indicates a sharply decreasing hazard rate. Because the time scale is time since energized, this implies that failures are more likely right after a radar is energized. It seems that on–off cycling has a harmful effect on the failure rate. As suggested by this article, however, one must be suspicious of decreasing hazard rates in the face of potential heterogeneity.

The presence of heterogeneity among units can be examined using the generalized residuals. The uncensored $\hat{F}(t_{ij})$'s, if there is no heterogeneity, should not differ across units. An analysis of variance on these $\hat{F}(t_{ij})$'s provided an F statistic with a value of 39.0 on 63 and 1,391 df, providing strong evidence that heterogeneity among units is present. This raises the possibility that the apparent decreasing hazard on time since energized is at least partially due to heterogeneity. To see whether a gamma mixing distribution was a reasonable way to model the heterogeneity, a Weibull model with the preceding explanatory variables and intercept dummies for individual radars was estimated. Only radars with at least 10 failures were included. These 53 $\hat{\lambda}_i$'s were treated as observations from a gamma distribution, and its mean and variance were estimated ($\hat{\mu} = .11$, $\hat{\sigma}^2 = .0018$). Figure 2 provides a histogram of the estimated $\hat{\lambda}_i$'s as well as a histogram of the predicted number of $\hat{\lambda}_i$'s assuming they follow a gamma distribution. A Pearson chi-squared test ($\sum_i (O_i - E_i)^2/E_i$) obtains a value of about 20 on 16 df. A gamma mixing distribution, it appears, is not an unreasonable assumption for the λ_i's.

Next, the gamma mixture of Weibull's model was estimated. Table 4 presents the parameter estimates. Whereas $\hat{\sigma}^2$ is significantly different from 0 using either a likelihood ratio or a Wald test, the estimate of $\hat{\gamma} = .772$ is quite similar to the estimate based on

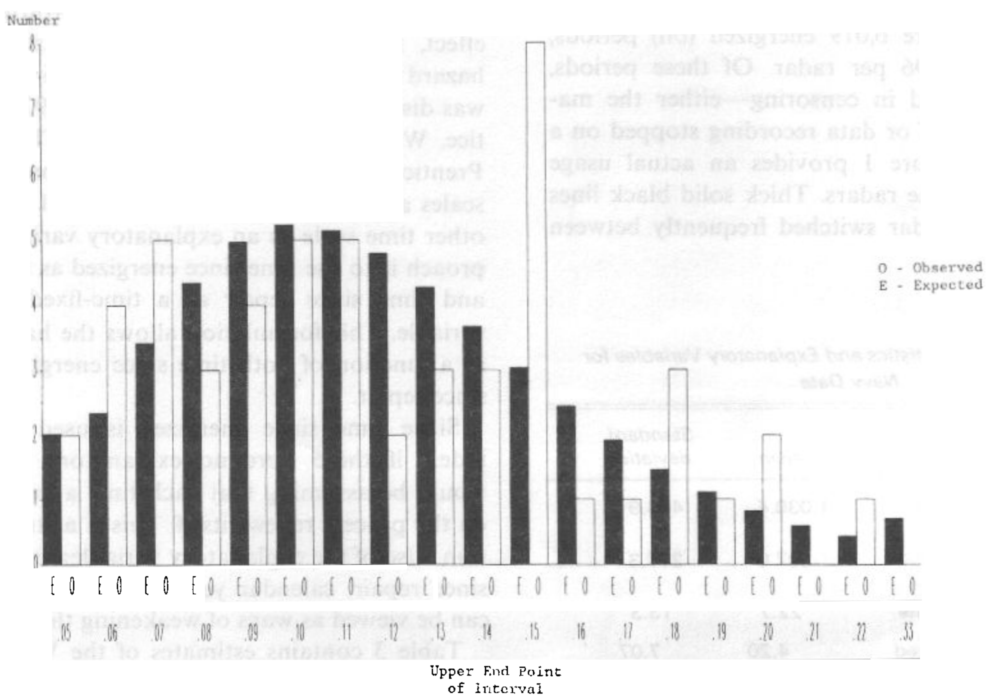

Figure 2. Histogram of Estimated $\hat{\lambda}_i$'s and the Predicted Number of $\hat{\lambda}_i$'s Under a Gamma Distribution. Estimated $\hat{\lambda}_i$'s are based on a Weibull regression with explanatory variables used in Table 3 and intercept dummies for the 53 radars with at least 10 failures. These $\hat{\lambda}_i$'s were then treated as a sample from a gamma distribution and $\hat{\mu} = .11$ and $\hat{\sigma}^2 = .0018$ estimated. These estimates provide the basis for the predicted number of $\hat{\lambda}_i$'s.

Table 4. Maximum Likelihood Estimates for the Weibull Mixture

Variable	Parameter	Standard error
μ	.112	.011
σ^2	.00073	.00030
γ	.772	.015
Days since repair	−.0037	.0022
First "on" since repair	.223	.078
On–off cycles per day	.046	.036
Calendar year	−.149	.044
Cumulative failures	.0026	.0039
Log-likelihood	−5,431.7	

the Weibull regression model. There appears to be a decreasing hazard on the time-since-energized, as well as heterogeneity across units. The other parameter estimates change somewhat from their previous values. In particular, only the first "on" since repair and calendar year remains significant at the .05 level. Perhaps the decreasing hazard on days since repair in the simple Weibull regression was partially due to heterogeneity.

An important question is whether the estimated parameters $\hat{\gamma}$ and $\hat{\sigma}^2$ are correlated. Both parameter estimates are significantly different from, respectively, 1.0 and .0; however, a high positive correlation might indicate that it is difficult to distinguish between heterogeneity and a decreasing hazard. A high correlation would indicate a ridge on the likelihood surface and values of γ less than 1.0 and σ^2 small would be supported by the data about as well as values of $\gamma = 1.0$ and σ^2 large. The interpretations for the two views, however, are quite different. In fact, the estimated correlation is −.05, which is close to 0 and negative.

The overall fit of the gamma mixture of Weibulls can be examined using a Kaplan–Meier estimate of the generalized residuals—that is, a Kaplan–Meier estimate treating the $(\hat{F}(t_{ij} \mid \hat{\lambda}_i), d_i)$'s as data from a distribution. Figure 3 displays this estimate. Recall that a line connecting (1, 0) and (0, 1) indicates a reasonable fit, and it appears that the fit of the Weibull mixture is adequate.

To test the proportional-hazard-rate assumption, we stratify the sample based on the value of each explanatory variable in turn. The proportional-hazard-rate assumption is valid if the hazard functions in each subsample are proportional as functions of time; that is, they have approximately the same value of the shape parameter, γ. Table 5 reports the results of this test. Since for each stratification the estimated γ's plus or minus their standard error overlap, the proportional-hazard-rate assumption receives some empirical support.

An important practical application of the model, given that it has been empirically validated, is to interpret $\hat{\gamma}$. Qualitatively, on–off cycling appears harmful because each time the system is energized, there is a high initial chance of failure. The effect of a decreasing hazard on time-since-energized can be

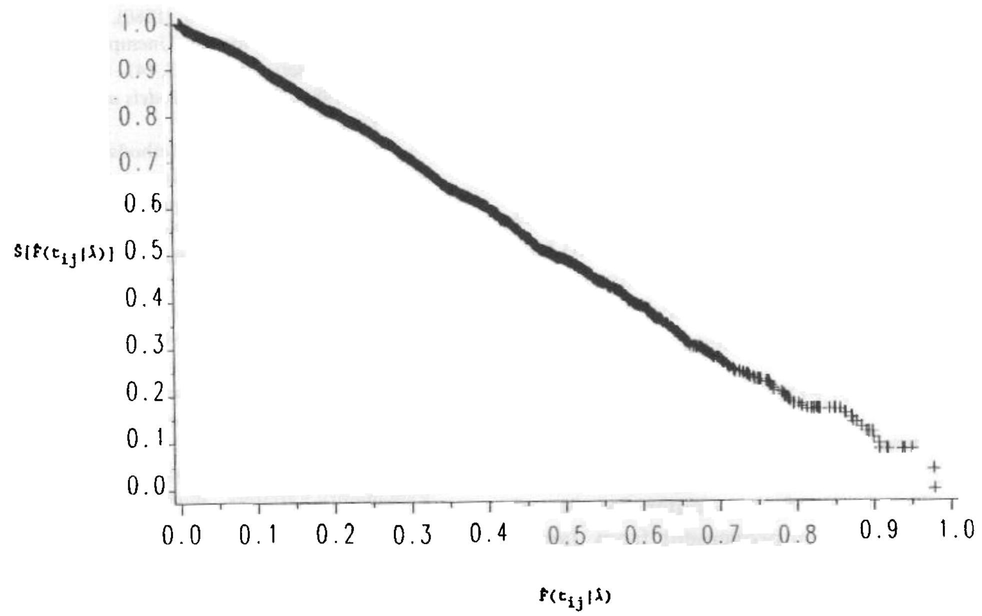

Figure 3. Kaplan–Meier Survival Curve of the Generalized Residuals Under the Weibull Mixture Model. If the model fits, the generalized residuals, $(\hat{F}(t_{ij}|\hat{\lambda}_i), d_i)$'s, should behave like a random sample from a uniform (0, 1) distribution. Therefore, the Kaplan–Meier estimate should follow a line connecting (1, 0) to (0, 1) if the Weibull mixture model fits the data.

Table 5. Maximum Likelihood Estimates by Strata

Variable	Value range	γ	Standard error	Sample size
Days since repair	<10	.759	.02	3,910
	>10	.786	.03	2,109
First "on" since repair	0	.757	.02	4,135
	1	.762	.02	1,884
On–off cycles per day	<1	.775	.02	4,250
	>1	.739	.03	1,769
Calendar year	80–82	.758	.02	3,380
	83–85	.773	.03	2,339
Cumulative failures	<20	.755	.02	3,420
	>20	.779	.02	2,599

quantified by use of the following simple scenario. Suppose that during a four-week period the radar needs to be continuously on, except for 28 hours of preventive maintenance (PM) performed while the radar is off. Suppose also that the machine was repaired just prior to this four-week period. The year is 1985, and the machine has had 10 cumulative failures. Two PM programs are considered; once every day for an hour or once every other day for two hours. The probability of no failure over the four-week period is .24 for daily PM and .29 for bidaily PM. The difference in estimated probabilities, althought slight, suggests that bidaily PM is preferred. In general, it is probably worthwhile for the Navy to turn this radar on and off less often where possible.

7. CONCLUSIONS

Heterogeneity across repairable units can cause an apparent decreasing hazard rate. The distinction between heterogeneity and a true decreasing hazard rate is important, because quite different conclusions follow regarding the risk of failure following repair. To sort out these separate effects, a statistical model was developed that allows for both. The model is applicable for data where a collection of units generates several failure times and where the failure rates may vary across units. The model was applied to two data sets where, respectively, a decreasing hazard rate was solely due to heterogeneity and both a decreasing hazard rate and heterogeneity were present.

ACKNOWLEDGMENTS

This research was sponsored by the Center for Naval Analyses. The findings, however, are ours and do not reflect the views of the Center for Naval Analyses. Suggestions for improvement from James Hodges, Lee Lillard, Donald Muraco, William Meeker, and two anonymous referees are gratefully acknowledged.

[*Received February 1987. Revised April 1988.*]

REFERENCES

Ascher, H. E., and Feingold, H. (1984), *Repairable Systems Reliability: Modeling, Inference, Misconceptions, and Their Causes*, New York: Marcel Dekker.

Chernoff, H. (1954), "On the Distribution of the Likelihood Ratio," *Annals of Mathematical Statistics*, 25, 573–578.

Cox, D. R., and Oakes, D. (1984), *Analysis of Survival Data*, New York: Chapman & Hall.

Crowder, M. (1985), "A Distributional Model for Repeated Failure Time Measurements," *Journal of the Royal Statistical Society*, Ser. B, 47, 447–452.

Dennis, J., and Schnabel, R. (1983), *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Englewood Cliffs, NJ: Prentice-Hall.

Dubey, S. (1968), "A Compound Weibull Distribution," *Naval Research Logistics Quarterly*, 15, 179–188.

Elber, C., and Ridder, G. (1982), "True and Spurious Duration Dependence: The Identifiability of the Proportional Hazard Model," *Review of Economic Studies*, 51, 402–411.

Farewell, V. T., and Cox, D. R. (1979), "A Note on Multiple Time Scales in Life Testing," *Applied Statistics*, 28, 73–74.

Heckman, J., and Singer, B. (1984), "The Identifiability of the Proportional Hazard Model," *Review of Economic Studies*, 51, 231–241.

Johnson, N., and Kotz, S. (1972), *Distributions in Statistics: Continuous Multivariate Distributions*, New York: John Wiley.

Kalbfleisch, J., and Prentice, R. (1980), *The Statistical Analysis of Failure Time Data*, New York: John Wiley.

Kay, R. (1977), "Proportional Hazard Regression Models and the Analysis of Censored Survival Data," *Applied Statistics*, 26, 227–237.

Kujawski, G., and Rypka, E. (1978), "Effects of On–Off Cycling on Equipment Reliability," in *Proceedings of the 1978 Annual Reliability and Maintainability Symposium*, New York: Institute of Electrical and Electronic Engineers, pp. 225–230.

Lagakos, S. (1979), "General Right Censoring and Its Impact on the Analysis of Survival Data," *Biometrics*, 35, 139–156.

Lancaster, T. (1979), "Econometric Models for the Duration of Unemployment," *Econometrica*, 47, 939–956.

Lancaster, T., and Nickell, S. (1980), "The Analysis of Reemployment Probabilities for the Unemployed," *Journal of the Royal Statistical Society*, Ser. A, 143, 141–152.

Lawless, J. (1982), *Statistical Models and Methods for Lifetime Data*, New York: John Wiley.

——— (1987), "Regression Methods for Poisson Process Data," *Journal of the American Statistical Association*, 82, 808–815.

Lee, L. (1980a), "Testing Adequacy of the Weibull and Log-Linear Rate Models for a Poisson Process," *Technometrics*, 22, 195–199.

——— (1980b), "Comparing Rates of Several Independent Weibull Processes," *Technometrics*, 22, 427–430.

Manton, K., Stallard, E., and Vaupel, J. (1986), "Alternative Models for the Heterogeneity of Mortality Risks Among the Aged," *Journal of the American Statistical Association*, 81, 635–644.

Moran, P. (1971), "Maximum Likelihood Estimation in Nonstandard Conditions," *Proceedings of the Cambridge Philosophical Society*, 70, 441–450.

Prentice, R. L. (1973), "Exponential Survivals With Censoring and Explanatory Variables," *Biometrika*, 60, 279–288.

Prentice, R., Williams, B., and Peterson, A. (1981), "On the Regression Analysis of Multivariate Failure Time Data," *Biometrika*, 68, 373–380.

Proschan, F. (1963), "Theoretical Explanation of Observed Decreasing Failure Rates," *Technometrics*, 5, 375–383.

Wild, C. J. (1983), "Failure Time Models With Matched Data," *Biometrika*, 70, 633–642.